Panel Unit Root Tests in the Presence of Cross-Sectional Dependency and Heterogeneity[1]

Yoosoon Chang
and
Wonho Song
Department of Economics
Rice University

Abstract

An IV approach, using as instruments nonlinear transformations of the lagged levels, is explored to test for unit roots in panels with general dependency and heterogeneity across cross-sectional units. We allow not only for the cross-sectional dependencies of innovations, but also for the presence of cointegration across cross-sectional levels. Unbalanced panels and panels with differing individual shortrun dynamics and cross-sectionally related dynamics are also permitted. Panels with such cross-sectional dependencies and heterogeneities appear to be quite commonly observed in practical applications. Yet none of the currently available tests can be used to test for unit roots in such general panels. We also more carefully formulate the unit root hypothesis in panels. In particular, using order statistics we make it possible to test for and against the presence of unit roots in some of the individual units for a given panel. The individual IV t-ratios, which are the bases of our tests, are asymptotically normally distributed and cross-sectionally independent. Therefore, the critical values of the order statistics as well as the usual averaged statistic can be easily obtained from simple elementary probability computations. We show via a set of simulations that our tests work well, while other existing tests fail to perform properly. As an illustration, our tests are applied to some of the data sets that were used in earlier studies.

This version: January 28, 2002

JEL Classification: C12, C15, C32, C33.
Key words and phrases: Panels with cross-sectional dependency and heterogeneity, unit root test, cointegration, covariate, nonlinear instrument, order statistics.

[1] This paper was prepared for the presentation at the 2002 ASSA Meeting held in Atlanta. We would like to thank Roger Moon and Joon Park for helpful discussions and comments. Correspondence address to: Yoosoon Chang, Department of Economics - MS 22, Rice University, 6100 Main Street, Houston, TX 77005-1892, Tel: 713-348-2796, Fax: 713-348-5278, Email: yoosoon@rice.edu.

1. Introduction

Panel unit root tests have been one of the most active research area for the past several years. This is largely due to the availability of panel data with long time span, and the growing use of cross-country and cross-region data over time to test for many important economic inter-relationships, especially those involving convergencies/divergencies of various economic variables. The notable contributers in theoretical research on the subject include Levin, Lin and Chu (1997), Im, Pesaran and Shin (1997), Maddala and Wu (1999), Choi (2001a) and Chang (1999, 2000). There have been numerous related empirical researches as well. Examples include MacDonald (1996), Oh (1996) and Papell (1997) just to name a few. The papers by Banerjee (1999), Phillips and Moon (1999) and Baltagi and Kao (2000) provide extensive surveys on the recent developments on the testing for unit roots in panels. See also Choi (2001b) and Phillips and Sul (2001) for some related work in this line of research.

In this paper, we consider an IV approach, using as instruments nonlinear transformations of the lagged levels. The idea was explored earlier by Chang (2000) to develop the tests that can be used for panels with cross-sectional dependencies in innovations of unknown form. Our work extends the approach by Chang (2000) in several important directions. First, we allow for the presence of cointegration across cross-sectional units. It appears that there is a high potential for such possibilities in many panels of practical interests. Yet, none of the existing tests, including the one developed by Chang (2000), is not applicable for such panels. Second, our tests are based on the models augmented by cross-sectional dynamics and other covariates. As demonstrated by Hansen (1995) and Chang, Sickles and Song (2000), the inclusion of covariates can dramatically increase the power of the tests. Third, we formulate the panel unit root hypothesis more carefully. In particular, we consider the null and alternative hypotheses that some, not all, of the cross-sectional units have unit roots. Such hypotheses are often more relevant for practical applications.

The presence of cointegration is dealt with simply by using an orthogonal set of functions as instrument generating functions. Chang (2000) considers the IV t-ratios based on the instruments generated by a single function for all cross-sectional units, and shows their asymptotic independence for panels with general cross-sectional dependency. However, as we demonstrate in the paper, the asymptotic independence of the IV t-ratios may be violated in the presence of cointegration across cross-sectional units, which would invalidate the tests by Chang (2000). It is shown in the paper that this difficulty can be resolved if we use the instruments generated by a set of functions that are orthogonal each other. If a set of orthogonal instrument generating functions are used, the resulting IV t-ratios become asymptotically independent in the presence of cointegration as well as the cross-correlation of innovations.

One of the main motivations to use panels to test for unit roots is to increase the power. An important possibility, however, has been overlooked here, i.e., the possibility of using covariates. The idea of using covariates to test for a unit root was first suggested by Hansen (1995), and its implementation using bootstrap was studied later by Chang, Sickles and Song (2001). They made it clear that there is a huge potential gain in power if covariates are appropriately chosen. Of course, the choice of proper covariates may be difficult in practical applications. In the panel context, however, some of potential covariates to account for the inter-relatedness of cross-sectional dynamics naturally come upfront. For instance, we may include the lagged differences of other cross-sections to allow for interactions in shortrun dynamics and the linear combinations of cross-sectional levels in the presence of cointegration.

Obviously, the power increase is not the only reason to test for unit roots in panels. We are often interested in testing for unit roots collectively for cross-sectional units included in a certain panel. In this case, it is necessary to formulate the hypotheses more carefully. In particular, we may want to test for and against the existence of the unit roots in not all, but only a fraction of cross-sectional units. Such formulation is, however, more appropriate to investigate important hypotheses such as purchasing power parity and growth convergence, among many others. The hypotheses can be tested more effectively using order statistics such as maximum and minimum of individual tests. As we show in the paper, the order statistics constructed from individual nonlinear IV t-ratios have limit distributions which are nuisance parameter free and given by simple functions of the standard normal. The critical values are thus easily derived from those of the standard normal distribution.

As should have now become obvious, our model is truly general. It allows for the cross-sectional dependency in both the longrun and the shortrun. We permit not only the cross-correlation of the innovations and/or cross-sectional dynamics in the shortrun, but also the comovements of the stochastic trends in the longrun. Our formulation of the hypotheses is also sharper and makes it possible to test for and against the partial existence of unit roots in panels. Yet our limit theories are all Gaussian and extremely simple to derive. All these flexibility and simplicity are due to the employment of the nonlinear IV methodology, or more specifically, the asymptotic independence and normality of the IV t-ratios. All other existing approaches do not offer such generality, assuming either cross-sectional independence that is unacceptable in most applications or a specific form of cross-sectional correlation structure that may be of only limited applicability.

We conduct a set of simulations to see whether our tests work properly in panels with usual cross-sectional and time dimensions, N and T. Though not extensively done, it seems clear that our tests perform reasonably well and are preferred to other existing tests. In particular, our average test, which is comparable to other existing tests, performs significantly better than other tests when N is large relative to T. It works quite well for T as small as 25 and N as big as 100. Not surprisingly, the order statistics to test for and against the presence of unit roots in a small fraction of cross-sectional units require relatively large T to work properly. For their reliable performance, we believe that T should be 500 or bigger. They are not much sensitive to the size of N. For the purpose of illustration, we apply our tests to examine the purchasing power parity conditions. It seems that the parity holds for some, but not all, countries considered in our study.

The rest of the paper is organized as follows. Section 2 specifies assumptions and background theory. The models and hypotheses are introduced, and some of the preliminary theories are included. Section 3 defines the test statistics for individual cross-sectional units and for panels, and develop their asymptotics. The results from simulation and empirical applications are summariazed in Section 4, and the concluding remarks follow in Section 5. The mathematical proofs are included in Appendix.

2. Assumptions and Background Theory

We consider a panel model generated as the following first order autoregressive regression:

$$y_{it} = \alpha_i y_{i,t-1} + u_{it}, \quad i = 1, \ldots, N; \; t = 1, \ldots, T_i. \tag{1}$$

As usual, the index i denotes individual cross-sectional units, such as individuals, households, industries or countries, and the index t denotes time periods. The number of time series observations T_i for each individual i may differ across cross-sectional units. Hence, unbalanced panels are allowed in our model.

2.1 Unit Root Hypotheses

We are interested in testing the unit root null hypothesis for the panel given in (1). More precisely, we consider the following sets of hypotheses.

Hypotheses (A) $\mathrm{H}_0 : \alpha_i = 1$ for all i *versus* $\mathrm{H}_1 : \alpha_i < 1$ for all i

Hypotheses (B) $\mathrm{H}_0 : \alpha_i = 1$ for all i *versus* $\mathrm{H}_1 : \alpha_i < 1$ for some i

Hypotheses (C) $\mathrm{H}_0 : \alpha_i = 1$ for some i *versus* $\mathrm{H}_1 : \alpha_i < 1$ for all i

Hypotheses (A) and (B) both include the same null hypothesis, which implies that the unit root is present in all individual units. However, their null hypothesis competes with different alternative hypotheses. It is tested in Hypotheses (A) against the hypothesis that all individual units are stationary, while in Hypotheses (B) the alternative is that there are some stationary individual units. On the contrary, the null hypothesis in Hypotheses (C) holds as long as the unit root exists in at least one individual unit, and is tested against the alternative hypothesis that all individual units are stationary. The alternative hypotheses in both Hypotheses (B) and (C) are negations of their null hypotheses. This is not the case for Hypotheses (A).

Virtually all the existing literature on panel unit root tests effectively looks at Hypotheses (A). Some recent work, including Im, Pesaran and Shin (1998) and Chang (2000), allow for heterogeneous panels, and formulate the null and alternative hypotheses as in Hypotheses (B). However, their use of average t-ratios can only be justified for the test of Hypotheses (A). To properly test Hypotheses (B), the minimum, instead of the average, of individual t-ratios should have been used. For the test of Hypotheses (B), the tests based on the minimum would clearly dominate those relying on the averages. In particular, the former is expected to have much larger power than the latter when only a small fraction of individual units are stationary.

Hypotheses (C) have never been considered in the literature, though they seem to be more relevant in many interesting empirical applications such as tests for purchasing power parities and growth convergences. Note that the rejection of H_0 in favor of H_1 in Hypotheses (C) directly implies that all (y_{it})'s are stationary, and therefore, purchasing power parities or growth convergences hold if we let (y_{it})'s be real exchange rates or differences in growth rates respectively. No test, however, is available to deal with Hypotheses (C) appropriately. Here we propose to use the maximum of individual t-ratios for the test of Hypotheses (C).

2.2 Shortrun Dynamics

We now completely specify the data generating process for our model introduced in (1). The initial values $(y_{10}, \ldots, y_{N0})'$ of $(y_{1t}, \ldots, y_{Nt})'$ do not affect our subsequent asymptotic analysis as long as they are stochastically bounded, and therefore we set them at zero for expositional brevity. We let $y_t = (y_{1t}, \ldots, y_{Nt})'$ and assume that there are $N - M$ cointegrating relationships in the unit root process (y_t), which are represented by the cointegrating vectors (c_j), $j = 1, \ldots, N - M$. The usual vector autoregression and error correction representation allows us to specify the shortrun dynamics of (y_t) as

$$\triangle y_{it} = \sum_{j=1}^{N} \sum_{k=1}^{P_{ij}} a_{ij} \triangle y_{j,t-k} + \sum_{j=1}^{N-M} b_{ij} c_j' y_{t-1} + \varepsilon_{it} \tag{2}$$

for each cross-sectional unit, where (ε_{it}) are white noise, $i = 1, \ldots, N$, and $\triangle$ is the difference operator.

Moreover, due to the Granger representation theorem, we may write $u_t = \triangle y_t$ as

$$u_t = \Pi(L)\varepsilon_t$$

where $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$, L is the lag operator, and $\Pi(z) = \sum_{k=0}^{\infty} \Pi_k z^k$ with $\Pi_0 = I$. We assume

Assumption 2.1 (Π_k) is 1-summable, and $\Pi(1)$ is of rank M for $M \leq N$.

It follows from the Beveridge-Nelson decomposition that

$$u_t = \Pi(1)\varepsilon_t + (\tilde{u}_{t-1} - \tilde{u}_t)$$

where

$$\tilde{u}_t = \tilde{\Pi}(L)\varepsilon_t$$

with $\tilde{\Pi}(z) = \sum_{k=0}^{\infty} \tilde{\Pi}_k z^k$ and $\tilde{\Pi}_k = \sum_{j=k+1}^{\infty} \Pi_j$. Consequently,

$$y_t = \Pi(1) \sum_{k=1}^{t} \varepsilon_k + (\tilde{u}_0 - \tilde{u}_t)$$

Note that $(\tilde{\Pi}_k)$ is absolutely summable due to the 1-summability of (Π_k) given in Assumption 2.1, and therefore, $(\tilde{u}_t)$ is well defined and stationary. Moreover, if $M < N$ and if C is an $N \times (N - M)$ matrix such that $C'\Pi(1) = 0$, then each column (c_j), $j = 1, \ldots, N - M$, of C represents a cointegrating vector for (y_t).

The data generating process for the innovations (ε_t) is assumed to satisfy the following assumption.

Assumption 2.2 (ε_t) is an iid $(0, \Sigma)$ sequence of random variables with $\mathbf{E}|\varepsilon_t|^\ell < \infty$ for some $\ell > 4$, and its distribution is absolutely continuous with respect to Lebesgue measure and has characteristic function φ such that $\lim_{s\to\infty} |s|^r \varphi(s) = 0$, for some $r > 0$.

Assumption 2.2 lays out the technical conditions that are required to invoke the asymptotic theories for the nonstationary nonlinear models developed by Park and Phillips (1999).

Our unit root tests at individual levels will be based on the regression

$$y_{it} = \alpha_i y_{i,t-1} + \sum_{k=1}^{P_i} \alpha_{ik} \triangle y_{i,t-k} + \sum_{k=1}^{Q_i} \beta'_{ik} w_{i,t-k} + \varepsilon_{it} \tag{3}$$

for $i = 1, \ldots, N$, where we interpret (w_{it}) as *covariates* added to the augmented Dickey-Fuller (ADF) regression for the i-th cross-sectional unit. It is important to note that the vector autoregression and error correction formulation of the cointegrated unit root panels in (2) suggests that we use such covariates. We may obviously rewrite (1) and (2) as (3) with several lagged differences of other cross-sections and linear combinations of lagged levels of all cross-sections as covariates. In the subsequent development of our theory, we will assume that the data generating process is given by (2) under Assumptions 2.1 and 2.2. This, however, is just for the expositional convenience. We may easily accommodate other covariates accounting for idiosyncratic characteristics of cross-sectional units, as long as they satisfy the conditions laid out in Chang, Sickles and Song (2001).

The unit root regression with covariates was first considered in Hansen (1995) and studied subsequently by Chang, Sickles and Song (2001). It was referred to by them as covariates augmented Dickey-Fuller (CADF) regression. Both Hansen (1995) and Chang, Sickles and Song (2001) show that using covariates offers a great potential in power gain for the test of a unit root. In many panels of interest, we naturally expect to have shortrun dynamics that are inter-related across different cross-sectional units, which would make it necessary to include the dynamics of others to properly model own dynamics. It is even necessary to take into consideration the longrun trends of other cross-sectional units in the presence of cointegration, since then error correction mechanism comes into play and the stochastic trends of other cross-sectional units would interfere with the own shortrun dynamics.

2.3 Basic Tools for Asymptotics

Here we introduce some basic theories that are needed to develop the asymptotics of our statistics. Define a stochastic processes U_T for ε_t as

$$U_T(r) = T^{-1/2} \sum_{t=1}^{[Tr]} u_t$$

on $[0, 1]$, where $[s]$ denotes the largest integer not exceeding s. The process $U_T(r)$ takes values in $D[0, 1]^N$, where $D[0, 1]$ is the space of cadlag functions on $[0, 1]$. Under Assumptions 2.1 and 2.2, an invariance principle holds for U_T, viz.

$$U_T \to_d U \tag{4}$$

as $T \to \infty$, where U is an N-dimensional vector Brownian motion with covariance matrix , where

$$= \Pi(1)\Sigma\Pi(1)'$$

Under Assumption 2.2, the covariance matrix is in general singular with rank M.

Our asymptotic theory involves the local time of Brownian motion, which we will introduce briefly below. The reader is referred to Park and Phillips (1999, 2001) and Chang, Park and Phillips (2001) and the references cited there for the concept of local time and its use in the asymptotics for nonlinear models with integrated time series. The local time L_i of U_i, for $i = 1, \ldots, N$, is defined by

$$L_i(t, s) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^t 1\{|U_i(r) - s| < \epsilon\} \, dr.$$

Roughly, the local time L_i measures the time that the Brownian motion U_i spends in the neighborhood of s, up to time t. It is well known that L_i is continuous in both t and s. For any local integrable function G on $\mathbf{R}$, we have an important formula

$$\int_0^t G(U_i(r)) \, dr = \int_{-\infty}^{\infty} G(s) L_i(t, s) \, ds \tag{5}$$

which is called the occupation time formula.

2.4 Instrument Generating Functions

We consider the IV estimation of the augmented autoregression (3). To deal with the cross-sectional dependency, we use the following instrument generated by a nonlinear function F_i

$$F_i(y_{i,t-1})$$

for the lagged level $y_{i,t-1}$ for each cross-sectional unit $i = 1, \ldots, N$. For the augmented regressors $x_{it} = (\triangle y_{i,t-1}, \ldots, \triangle y_{i,t-P_i}; w_{i,t-1}, \ldots, w_{i,t-Q_i})'$, we use the variables themselves as the instruments. Hence for the entire regressors $(y_{i,t-1}, x'_{it})'$, we use the instruments given by

$$(F_i(y_{i,t-1}), x'_{it})' \tag{6}$$

similarly as in Chang (2000).

The transformations (F_i) will be referred to as the *instrument generating functions* (IGF). We assume that

Assumption 2.3 Let (F_i) be regularly integrable and satisfy (a) $\int_{-\infty}^{\infty} xF_i(x)dx \neq 0$ for all i and (b) $\int_{\infty}^{\infty} F_i(x)F_j(x)dx = 0$ for all $i \neq j$.

The class of *regularly integrable* transformations was first introduced in Park and Phillips (1999), to which the reader is referred for details. They are just transformations on $\mathbf{R}$ satisfying some mild technical regularity conditions.

Assumption 2.3(a) needs to hold, since otherwise we would have *instrument failure* and the resulting IV estimator becomes inconsistent. It is analogous to the non-orthogonality (between the instruments and regressors) requirement for the validity of IV estimation in standard stationary regressions. See Chang (2000) for more detailed discussions. Assumption 2.4(b) is necessary to allow for the presence of cointegration. If cointegration is present, the procedure in Chang (2000) relying on the same IGF's for all cross-sectional units becomes invalid. This will be explained in detail later in the next section.

The Hermite functions of odd orders $k = 2i - 1$, $i = 1, \ldots, N$, satisfy all the conditions in Assumption 2.3, and therefore, can be used as the IGF's. The Hermite function G_k of order k, $k = 0, 1, 2, \ldots$, is defined as

$$G_k(x) = (2^k k! \sqrt{\pi})^{-1/2} H_k(x) e^{-x^2/2} \tag{7}$$

where H_k is the Hermite polynomial of order k given by

$$H_k(x) = (-1)^k e^{x^2} \frac{d^k}{dx^k} e^{-x^2}$$

The shapes of Hermite functions of orders $k = 1, 3, 5$ and 7 are given in Figure 1. It is well known that the class of Hermite functions introduced above forms an orthonormal basis for $L^2(\mathbf{R})$, i.e., the Hilbert space of square integrable functions on $\mathbf{R}$. We thus have

$$\int_{-\infty}^{\infty} G_j(x) G_k(x) dx = \delta_{jk}$$

for all j and k, where δ_{jk} is the kronecker delta.

It is necessary to scale the functions (G_k) properly before we use them as instrument generating functions. Scaling the instrument generating functions determines the shapes and degrees of integrability of the functions, and consequently it affects the finite sample performances of the tests in an important way. The issue of how to scale the IGF's is indeed an empirical issue. However, we will discuss it here since it is critical for the finite sample performances. For actual practical implementations, there is no best way to scale the functions, and thus one needs to rely on simulation results to decide how to scale the IGF's. For panel unit root tests, the required adjustment involves both relative and global scalings. The relative scaling is needed to balance the functions across cross-sectional units, and eventually to ensure that the orthogonality holds for the generated instruments. The global scaling is needed to determine right degree of integrability of the function based on the data variation in each cross-sectional unit.

For relative scaling of (G_k), we normalize the data instead of directly scaling the arguments of the functions. This is just for the expositional convenience. Of course one may adjust scales for (G_k) by multiplying constants to their arguments, and obtain the same results. Note that the unit root test is invariant with respect to scale transformation, so our subsequent scale adjustment of the data can be done without loss of any generality. We define

$$v_i^2 = \frac{1}{T_i} \sum_{t=1}^{T_i} y_{it}^2$$

for $i = 1, \ldots, N$. Moreover, to scale the entire panel we fix one cross-sectional unit as a base for the adjustment and call it a *scale numeraire*. Any cross-sectional unit, say $i = 1$, can serve as the scale numeraire. We then let

$$\kappa_i = \upsilon_1/\upsilon_i$$

and define for $i = 1, \ldots, N$

$$y^*_{it} = \kappa_i y_{it} \tag{8}$$

We may call (y^*_{it}) *scale adjusted*. The scale adjustment would therefore make the sample standard deviations same for all cross-sectional units. If the panels are balanced, i.e., $T_i = T_j$ for all i and j, and if all the cross-sectional units are of the same scale so that all κ_i's asymptotically approach to unity in probability, the scale adjustment would be unnecessary. In what follows we assume that the scale adjustment is already done for (y_{it}), and continue to use (y_{it}) in place of (y^*_{it}) for expositional brevity. Under this convention, we now define the IGF's (F_i) by

$$F_i = G_{2i-1}$$

for $i = 1, \ldots, N$.

Once the data are scale-adjusted cross-sectionally, the global scaling for all cross-sections can be done by using the scaling factor for the scale numeraire. Since the factor for the scale numeraire is used for all cross-sections, it seems reasonable to choose as the scale numeraire the unit with the largest number of observations. Therefore, we set T_1 to be the maximum of T_i's for unbalanced panels in our subsequent discussions. The issue of choosing the scaling factor for the numerarie is discussed further in the simulation section.

3. Test Statistics and Their Asymptotics

In this section, we explicitly define test statistics and derive their asymptotic theories. We first look at IV t-ratios for individual cross-sectional units, and derive their asymptotics. We then discuss how one may combine the individual IV t-ratios in formulating tests for the panel unit root hypotheses, specified earlier as Hypotheses (A) – (C), and subsequently develop the asymptotics for the resulting statistics.

3.1 Individual IV t-ratios and Their Asymptotics

We first define individual IV t-ratios explicitly. Let

$$y_i = \begin{pmatrix} y_{i,P_i+1} \\ \vdots \\ y_{i,T_i} \end{pmatrix}, \quad y_{\ell i} = \begin{pmatrix} y_{i,P_i} \\ \vdots \\ y_{i,T_i-1} \end{pmatrix}, \quad X_i = \begin{pmatrix} x'_{i,P_i+1} \\ \vdots \\ x'_{i,T_i} \end{pmatrix}, \quad \varepsilon_i = \begin{pmatrix} \varepsilon_{i,P_i+1} \\ \vdots \\ \varepsilon_{i,T_i} \end{pmatrix}$$

where $x'_{it} = (\triangle y_{i,t-1}, \ldots, \triangle y_{i,t-P_i}; w'_{i,t-1}, \ldots, w'_{i,t-Q_i})$. Then the augmented autoregression (3) can be written in matrix form as

$$y_i = y_{i\ell}\alpha_i + X_i\gamma_i + \varepsilon_i = Y_i\delta_i + \varepsilon_i \tag{9}$$

where $\gamma_i = (\alpha_{i1}, \ldots, \alpha_{iP_i}; \beta_{i1}, \ldots, \beta_{iQ_i})'$, $Y_i = (y_{\ell i}, X_i)$, and $\gamma_i = (\alpha_i, \gamma'_i)'$. For the regression (9), we consider the estimator $\hat{\delta}_i$ of δ_i given by

$$\hat{\delta}_i = \begin{pmatrix} \hat{\alpha}_i \\ \hat{\gamma}_i \end{pmatrix} = (M'_iY_i)^{-1}M'_iy_i = \begin{pmatrix} F_i(y_{i\ell})'y_{\ell i} & F_i(y_{i\ell})'X_i \\ X'_iy_{i\ell} & X'_iX_i \end{pmatrix}^{-1} \begin{pmatrix} F_i(y_{i\ell})'y_i \\ X'_iy_i \end{pmatrix} \tag{10}$$

where $M_i = (F_i(y_{i\ell}), X_i)$ with $F_i(y_{il}) = (F_i(y_{i,P_i}), \ldots, F_i(y_{i,T_i-1}))'$ The estimator $\hat{\delta}_i$ is thus defined to be the IV estimator using the instruments M_i.

The IV estimator $\hat{\alpha}_i$ for the AR coefficient α_i corresponds to the first element of $\hat{\delta}_i$ given in (10). Under the null, we have

$$\hat{\alpha}_i - 1 = B_{T_i}^{-1} A_{T_i} \tag{11}$$

where

$$\begin{aligned}
A_{T_i} &= F_i(y_{i\ell})'\varepsilon_i - F_i(y_{\ell i})'X_i(X_i'X_i)^{-1}X_i'\varepsilon_i \\
&= \sum_{t=1}^{T_i} F_i(y_{i,t-1})\varepsilon_{it} - \sum_{t=1}^{T_i} F_i(y_{i,t-1})x_{it}' \left(\sum_{t=1}^{T_i} x_{it}x_{it}' \right)^{-1} \sum_{t=1}^{T_i} x_{it}\varepsilon_{it} \\
B_{T_i} &= F_i(y_{i\ell})'y_{i\ell} - F_i(y_{\ell i})'X_i(X_i'X_i)^{-1}X_i'y_{i\ell} \\
&= \sum_{t=1}^{T_i} F_i(y_{i,t-1})y_{i,t-1} - \sum_{t=1}^{T_i} F_i(y_{i,t-1})x_{it}' \left(\sum_{t=1}^{T_i} x_{it}x_{it}' \right)^{-1} \sum_{t=1}^{T_i} x_{it}y_{i,t-1}
\end{aligned}$$

and the variance of A_{T_i} is given by

$$\sigma_i^2 \mathbf{E} C_{T_i}$$

under Assumption 2.2, where

$$\begin{aligned}
C_{T_i} &= F_i(y_{i\ell})'F_i(y_{i\ell}) - F_i(y_{\ell i})'X_i(X_i'X_i)^{-1}X_i'F_i(y_{i\ell}) \\
&= \sum_{t=1}^{T_i} F_i(y_{i,t-1})^2 - \sum_{t=1}^{T_i} F_i(y_{i,t-1})x_{it}' \left(\sum_{t=1}^{T_i} x_{it}x_{it}' \right)^{-1} \sum_{t=1}^{T_i} x_{it}F_i(y_{i,t-1}).
\end{aligned}$$

For testing the unit root hypothesis $\alpha_i = 1$ for each $i = 1, \ldots, N$, we construct the t-ratio statistic from the nonlinear IV estimator $\hat{\alpha}_i$ defined in (11). More specifically, we construct such IV t-ratio for testing for a unit root in (1) or (3) as

$$\tau_i = \frac{\hat{\alpha}_i - 1}{s(\hat{\alpha}_i)} \tag{12}$$

where $s(\hat{\alpha}_i)$ is the standard error of the IV estimator $\hat{\alpha}_i$ given by

$$s(\hat{\alpha}_i)^2 = \hat{\sigma}_i^2 B_{T_i}^{-2} C_{T_i} \tag{13}$$

The $\hat{\sigma}_i^2$ is the usual variance estimator given by $T_i^{-1} \sum_{t=1}^{T_i} \hat{\varepsilon}_{it}^2$, where $\hat{\varepsilon}_{it}$ is the fitted residual from the augmented regression (3), viz.

$$\hat{\varepsilon}_{it} = y_{it} - \hat{\alpha}_i y_{i,t-1} - \sum_{k=1}^{P_i} \hat{\alpha}_{i,k} \triangle y_{i,t-k} - \sum_{k=1}^{Q_i} \hat{\beta}_{i,k} w_{i,t-k} = y_{it} - \hat{\alpha}_i y_{i,t-1} - x_{it}'\hat{\gamma}_i.$$

It is natural in our context to use the IV estimate $(\hat{\alpha}_i, \hat{\gamma}_i)$ given in (10) to get the fitted residual $\hat{\varepsilon}_{it}$. However, we may obviously use any other estimator of (α_i, γ_i) as long as it yields a consistent estimate for the residual error variance.

The limit null distribution of the IV t-ratio τ_i for testing $\alpha_i = 1$ defined in (12) is derived easily from the asymptotics for nonlinear transformations of integrated processes established in Park and Phillips (1999, 2001) and Chang, Park and Phillips (2001) and is given in

Lemma 3.1 Under Assumptions 2.1 – 2.3, we have

$$\tau_i \ \to_d \ \mathbf{N}(0,1)$$

as $T_i \to \infty$ for all $i = 1, \ldots, N$.

The normality of the limiting null distribution of the IV t-ratio τ_i is a direct consequence of using the instrument $F_i(y_{i,t-1})$ which is a regularly integrable transformation of the lagged level $y_{i,t-1}$, an integrated process under the unit root hypothesis. Our limit theory here is thus fundamentally different from the usual unit root asymptotics. This is due to the local time asymptotics and mixed normality of the sample moment $\sum_{t=1}^{T_i} F_i(y_{i,t-1})\varepsilon_{it}$ and the asymptotic orthogonalities between the instrument $F_i(y_{i,t-1})$ and the augmented variables $(\triangle y_{i,t-1}, \ldots, \triangle y_{i,t-P_i}; w_{i,t-1}, \ldots, w_{i,t-Q_i})$ which are all stationary. The nonlinearity of the instrument is therefore essential for our Gaussian limit theory. Moreover, the limit standard normal distributions are independent across cross-sectional units $i = 1, \ldots, N$, as we show in the next section.

We now consider the limit behavior of our IV t-ratio statistics under the alternative of stationarity to discuss the consistency of the test. Note that under the alternative, i.e., $\alpha_i = \alpha_{i0} < 1$, our IV t-ratio τ_i given in (12) can be expressed as

$$\tau_i = \tau_i(\alpha_{i0}) + \frac{\sqrt{T_i}(\alpha_{i0} - 1)}{\sqrt{T_i}\, s(\hat{\alpha}_i)} \tag{14}$$

where $s(\hat{\alpha}_i)$ is defined in (13) and

$$\tau_i(\alpha_{i0}) = \frac{\hat{\alpha}_i - \alpha_{i0}}{s(\hat{\alpha}_i)} \tag{15}$$

which is the IV t-ratio statistics for testing $\alpha_i = \alpha_{i0} < 1$. Under the alternative, we may expect that $\tau_i(\alpha_{i0}) \to_d \mathbf{N}(0,1)$ if the usual mixing conditions for (y_{it}) are assumed to hold. Moreover, if we let $B_{i0} = \operatorname{plim}_{T_i\to\infty} T_i^{-1} B_{T_i}$ and $C_{i0} = \operatorname{plim}_{T_i\to\infty} T_i^{-1} C_{T_i}$ exist under suitable mixing conditions for (y_{it}), then the second term in the right hand side of equation (14) diverges to $-\infty$ at the rate of $\sqrt{T_i}$. This is because

$$\sqrt{T_i}(\alpha_{i0} - 1) \to -\infty \quad \text{and} \quad \sqrt{T_i}\, s(\hat{\alpha}_i) \to_p \nu_i$$

where $\nu_i^2 = \sigma_i^2 B_{i0}^{-2} C_{i0} > 0$. Hence, the IV t-ratio τ_i diverges at the $\sqrt{T_i}$-rate under the alternative of stationarity, just as in the case of the usual OLS-based t-type unit root tests such as the augmented Dickey-Fuller test.

3.2 Test Statistics for Panels and Their Asymptotics

For the tests of Hypotheses (A) – (C), we let τ_i be the IV t-ratio for the i-th cross-sectional unit, and define

$$S = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i \tag{16}$$

$$S_{\min} = \min_{1\le i\le N} \tau_i \tag{17}$$

$$S_{\max} = \max_{1\le i\le N} \tau_i \tag{18}$$

The averaged statistic S is proposed for the test of Hypotheses (A), and comparable to other existing tests. The minimum statistic $S_{\min}$ is more appropriate for the test of Hypotheses (B). The average statistic S can also be used to test for Hypotheses (B), but the test based on $S_{\min}$ would have more power, especially when only a small fraction of cross-sectional units are stationary under the alternative hypothesis. The maximum statistic $S_{\max}$ can be used to test Hypotheses (C). Obviously, the average statistic S cannot be used to test for Hypotheses (C), since it would have incorrect size.

Let M be $0 \leq M \leq N$ and define

$$T_{\min} = \min_{1 \leq i \leq N} T_i, \quad T_{\max} = \max_{1 \leq i \leq N} T_i$$

We assume

Assumption 3.1 Let $\alpha_i = 1$ for $1 \leq i \leq M$, and set $M = 0$ if $\alpha_i < 1$ for all $1 \leq i \leq N$.

Assumption 3.2 Assume

$$T_{\min} \to \infty, \quad T_{\max}/T_{\min}^2 \to 0$$

which will simply be signified by $T \to \infty$ in our subsequent asymptotics.

Assumption 3.1 implies that there are M cross-sectional units having unit roots.[2] Assumption 3.2 gives the premier for our asymptotics. Our asymptotics are based on T-asymptotics and require that the time spans for all cross-sectional units be large for our asymptotics to work. However, we allow for unbalanced panels and they only need to be balanced asymptotically. Our conditions here are fairly weak, and we may therefore expect them to hold widely.

We have

Lemma 3.2 Under Assumptions 2.1 – 2.3 and 3.1 – 3.2, the results in Lemma 3.1 hold jointly for all $i = 1, \ldots, M$ independently across $i = 1, \ldots, M$.

The asymptotic independence of τ_i's are crucial for the subsequent development of our theory. Note that here we allow for the presence of cointegration as well as the unknown form of cross-sectional dependencies in innovations. We now explain the reason why we may expect their asymptotic independence even under such general cross-sectional dependencies. Assume for simplicity that the panels are balanced, i.e., $T_i = T$ for all i. As shown in Chang, Park and Phillips (2001), we have

$$\begin{aligned}
\frac{1}{\sqrt[4]{T}} \sum_{t=1}^{T} F_i(y_{i,t-1})\varepsilon_{it} \quad &\approx_d \quad \sqrt[4]{T} \int_0^1 F_i(\sqrt{T}U_i)dV_i \\
\frac{1}{\sqrt[4]{T}} \sum_{t=1}^{T} F_j(y_{j,t-1})\varepsilon_{jt} \quad &\approx_d \quad \sqrt[4]{T} \int_0^1 F_i(\sqrt{T}U_j)dV_j
\end{aligned}$$

which become independent if and only if their quadratic covariation

$$\sigma_{ij}\sqrt{T} \int_0^1 F_i(\sqrt{T}U_i(r))F_j(\sqrt{T}U_j(r))dr \to_{a.s.} 0 \tag{19}$$

as $T \to \infty$, where σ_{ij} denotes the covariance between V_i and V_j representing the limit Brownian motions of (ε_{it}) and (ε_{jt}) respectively.

It is indeed well known that

$$\int_0^1 F_i(\sqrt{T}U_i(r))F_j(\sqrt{T}U_j(r))dr = \log T/T \tag{20}$$

[2] We defined earlier M to be the number of independent unit roots, net of the number of cointegrating relationships. The presence of cointegration, however, no longer comes into our asymptotics, due to the orthogonality of the set of IGF's. We therefore use here M to denote the total number of unit roots.

for any Brownian motions U_i and U_j so long as they are not degenerate, and this implies that the condition (19) holds even when $\sigma_{ij} \neq 0$. Chang (2000) uses this result to develop the unit root tests for panels with cross-sectionally correlated innovations. However, (20) does not hold in the presence of cointegration between (y_{it}) and (y_{jt}). In this case, their limiting Brownian motions U_i and U_j become degenerate. If the cointegrating relationship is given by the unit coefficient, for instance, then we would have $U_i = U_j$, and therefore,

$$\begin{aligned} \sqrt{T}\int_0^1 F_i(\sqrt{T}U_i(r))F_j(\sqrt{T}U_j(r))dr &= \sqrt{T}\int_0^1 (F_iF_j)(\sqrt{T}U_i(r))dr \\ &= \sqrt{T}\int_{-\infty}^{\infty} (F_iF_j)(\sqrt{T}s)L_i(1,s)ds \\ &= \int_{-\infty}^{\infty} (F_iF_j)(s)L_i(1,s/\sqrt{T})ds \\ &= \left(\int_{-\infty}^{\infty} (F_iF_j)(s)ds\right) L_i(1,0) + o_{a.s.}(1) \end{aligned}$$

by the occupation times formula (5), change of variables and the continuity of $L(1,\cdot)$. The asymptotic independence of τ_i and τ_j generally breaks down, and holds only when F_i and F_j are orthogonal. This is the reason why the method by Chang (2000) becomes invalid in the presence of cointegration. We use an orthogonal set of IGF's to attain the asymptotic independence here.

The asymptotic theories for the statistics S, $S_{\min}$ and $S_{\max}$ may be easily derived from Lemma 3.2. We now let Φ be the distribution function for the standard normal distribution, and let λ be the size of the tests. For a given size λ, we define $x_M(\lambda)$ (with $x_1(\lambda) = x(\lambda)$) and $y_N(\lambda)$ by

$$\Phi(x_M(\lambda))^M = \lambda, \quad (1 - \Phi(y_N(\lambda)))^N = 1 - \lambda$$

These provide the critical values of the statistics $S, S_{\min}$ and $S_{\max}$ for the tests of Hypotheses (A) – (C). The following table shows the tests and critical values that can be used to test each of Hypotheses (A) – (C).

Hypotheses	Test Statistics	Critical Values
Hypotheses (A)	S	$x(\lambda)$
Hypotheses (B)	S $S_{\min}$	$x(\lambda)$ $y_N(\lambda)$
Hypotheses (C)	$S_{\max}$ $S_{\max}$	$x(\lambda)$ $x_M(\lambda)$

The critical values $x_M(\lambda)$ and $y_N(\lambda)$ for sizes $\lambda = 1\%, 5\%$ and 10% are tabulated in Tables A and B for the cases of $M, N = 2, \ldots, 121$.

The following theorem summarizes the asymptotic behaviors of $S, S_{\min}$ and $S_{\max}$.

Theorem 3.2 Let Assumptions 2.1 – 2.3 and 3.1 – 3.2 hold. If $M = N$,

$$\begin{aligned} \lim_{T\to\infty} \mathbf{P}\{S \leq x(\lambda)\} &= \lambda \\ \lim_{T\to\infty} \mathbf{P}\{S_{\min} \leq y_N(\lambda)\} &= \lambda \end{aligned}$$

If $1 \leq M \leq N$, then

$$\lim_{T\to\infty} \mathbf{P}\{S_{\max} \leq x_M(\lambda)\} = \lambda, \quad \lim_{T\to\infty} \mathbf{P}\{S_{\max} \leq x(\lambda)\} \leq \lambda$$

On the other hand, $S, S_{\max} \to_p -\infty$ if $M = 0$, and $S_{\min} \to_p -\infty$ if $M < N$.

Theorem 3.2 implies that all our tests have the prescribed asymptotic sizes. The tests using statistics S and $S_{\min}$ with critical values $x(\lambda)$ and $y_N(\lambda)$, respectively, have the exact size λ asymptotically under the null hypotheses in Hypotheses (A) and (B). However, the null hypothesis in Hypotheses (C) is composite, and the rejection probabilities of the test relying $S_{\max}$ with critical values $x(\lambda)$ may not be exactly λ even asymptotically. The size λ in this case is the maximum rejection probabilities that may result in under the null hypothesis. If we have a prior belief about the number M of cross-sectional units with a unit root, then we may more precisely formulate Hypotheses (C) as

Hypotheses (C′) $\quad \mathrm{H}_0 : \alpha_i = 1$ for M i's *versus* $\mathrm{H}_1 : \alpha_i < 1$ for all i

In this case, the test $S_{\max}$ will have exact asymptotic size since we can use the critical value $x_M(\lambda)$. Moreover, the null and the alternative are now further away from each other, and therefore we may expect $S_{\max}$ to have higher discriminating power for Hypotheses (C′).

Theorem 3.2 also shows that all our tests are consistent for Hypotheses (A) – (C).

4. Simulations and Empirical Illustrations

In this section, we conduct a set of simulations to investigate finite sample performances of the newly proposed panel unit root tests based on the covariates augmented nonlinear IV t-ratios for testing the unit root in individual units. The average (S), maximum ($S_{\max}$) and minimum ($S_{\min}$) statistics constructed from the covariates augmented nonlinear IV t-ratios are carefully examined and compared to two of the existing average tests: (a) the test (S_N) based on the standard nonlinear IV t-ratios by Chang (2000) and (b) the t-bar test by Im, Pesaran and Shin (1997) based on the usual OLS t-ratios with mean and variance adjustments.

For the simulations, we consider a simple heterogeneous, dependent and cross-sectionally cointegrated panel model. More explicitly, we consider the model (1) with (y_{it}) generated by the following DGP

$$u_{it} = \xi_t + \theta\xi_{t-1} + \eta_{it} - \eta_{i,t-1}, \quad i = 1, \ldots, N;\ t = 1, \ldots, T \tag{21}$$

where ξ_t is the scalar common stochastic trend and $\eta_t = (\eta_{1t}, \ldots, \eta_{Nt})'$ an N-dimensional innovation vector. The processes ξ_t and η_t are independent and drawn respectively from $ii\mathbf{N}(0,1)$ and $ii\mathbf{N}(0,V)$. Here the cross-sectional cointegrations are driven by the common stochastic trend ξ_t which generates $(N-1)$ cointegrating relationships among N individual units. There exists a cointegrating relationship between any pair of y_{it} and y_{jt} and all of the cointegrating relationships are characterized by the common cointegrating vector $(1,-1)'$. Notice that the common trend ξ_t enters as a moving average, thereby generating not only individual serial correlations but also cross-sectionally related dynamics. The cross-sectional dependencies are introduced by the dependencies of the innovations η_{it}'s through their covariance matrix V which is unrestricted except for being symmetric and nonsingular.

The moving average coefficient θ of the common trend is drawn randomly from Uniform[-0.2,0.2]. The parameters of $(N \times N)$ covariance matrix V of the innovations η_t are also drawn randomly. For the random generation of V we need to ensure that V is a symmetric positive definite matrix and avoid the near singularity problem. To do this we follow the steps outlined in Chang (1999) and provide them here for convenience:

(1) Generate an $(N \times N)$ matrix M from Uniform[0,1].
(2) Construct from M an orthogonal matrix $H = M(M'M)^{-1/2}$.
(3) Generate a set of N eigenvalues, $\lambda_1, \ldots, \lambda_N$. Let $\lambda_1 = r > 0$ and $\lambda_N = 1$ and draw $\lambda_2, \ldots, \lambda_{N-1}$ from Uniform$[r,1]$.
(4) Form a diagonal matrix Λ with $(\lambda_1, \ldots, \lambda_N)$ on the diagonal.
(5) Construct the covariance matrix V as a spectral representation $V = H\Lambda H'$.

Constructed as such the covariance matrix V will be symmetric and nonsingular with eigenvalues taking values from r to 1. The parameter r determines the ratio of the minimum to the maximum eigenvalue and thus provides a measure for degree of correlatedness among the components components. The covariance matrix V becomes singular as r tends to zero, and becomes spherical as r approaches to 1. For the simulations, we set r at 0.1.

The error process $\triangle y_{it} = u_{it}$ generated as in (21) can be specified as (2) using the usual VAR and ECM representations (possibly with infinite number of lagged differences), and naturally we may base our unit root testing on the covariates augmented regression given in (3). For the practical implementation, we need to choose the covariates w_{it} as well as the order P_i for the own lagged differences. Here we choose the covariates for each cross-section by using a practical scheme which simultaneously picks a best combination of its own lagged differences, other cross-sections' first lagged differences and finally from the lagged cointegration errors. The order for the lagged differences is based on the AIC rule with the maximum lag order 3. The covariates are selected by choosing those with higher correlations with the error. This is to make the effective error have smallest variance, which will in turn lead to largest power gain from the inclusion of the covariates.

To accommodate cross-sectionally cointegrated and dependent data, we use a set of N orthogonal instrument generating functions $F_i = G_k,\ k = 2i-1, i = 1, \ldots, N$ based on the Hermite functions (G_k) defined in (7). The relative scaling of the IGF's is done simply by fixing $i = 1$ as the scale numeraire and by transforming remaining y_{it}'s, $i = 2, \ldots, N$ by multiplying them with the factors κ_i's defined above (8). For the global scaling, we first scale the the numeraire in the data-dependent way suggested in Chang (2000). That is, for $i = 1$, we generate the instrument for the lagged level $y_{1,t-1}$ by $F_1(cy_{1,t-1})$ The scale factor c is determined by $c = KT^{-1/2}s^{-1}(\triangle y_{1t})$ with $s^2(\triangle y_{1t}) = T^{-1}\sum_{t=1}^{T}(\triangle y_{1t})^2$ where K is a constant.[3] The factor c is in particular inversely proportional to the sample standard error of $\triangle y_{1t} = u_{1t}$, but due to the relative scaling, all y_{it}'s have the same variation. Thus we use the same factor c to scale the arguments in F_i's and use $F_i(cy_{i,t-1})$ as the instrument for $y_{i,t-1}$ for all $i = 2, \ldots, N$. This section is yet to be completed.

5. Conclusions

This paper extends the existing methodologies for panel unit root tests in three important directions. First, we allow for dependencies across individual cross-sections at both shortrun and longrun levels. We allow for inter-relatedness of cross-sectional shortrun dynamics and the presence of longrun relationships in cross-sectional levels. Many panels of practical interests seem to have such complicated cross-sectional dependencies. Second, our theory permits the use of covariates to increase the power. Covariates may naturally include the terms to account for cross-sectional dependencies, but others to control idiosyncrasies of individual cross-sectional units may also well be potential candidates. If properly chosen, the inclusion of covariates would substantially improve the power of the test, as demonstrated earlier by several authors. Third, we re-examined the formulation of the unit root

[3] The value of K is fixed at 5 for all $i = 1, \ldots, N$ and for all combinations of N and T considered here.

hypothesis in panels, and propose to look at the null and alternative hypotheses that only a fraction of cross-sectional units have unit roots. Such formulations are more appropriate for some of the most commonly investigated panel models such as purchasing power parity and growth convergence.

The tests developed in the paper are valid for very general panels. They allow not only for unknown forms of cross-sectional dependencies at several different levels, but also for various kinds of heterogeneities such as unbalancedness, differing dynamics and other idiosyncratic characteristics for individual units. These indeed appear to be the common characteristics of many panels used in empirical studies. Neverthless, none of the currently available tests is applicable for such panels. In spite of their applicability, our tests are truly simple to implement. The relevant statistical theories are quite simple and all Gaussian, and the critical values are given by either standard normal or its simple functionals.

Appendix: Mathematical Proofs

Proof of Lemma 3.1 The asymptotics of the following sample moments involving integrable transformations of unit root process follows directly from Park and Phillips (1999,2001) as

$$\begin{aligned} T_i^{-1/4}\sum_{t=1}^{T_i} F_i(y_{i,t-1})\varepsilon_{it} &\to_d \quad \mathbf{MN}\left(0, \sigma_i^2 L_i(1,0)\int_{-\infty}^{\infty} F_i(s)^2 ds\right) \\ T_i^{-1/2}\sum_{t=1}^{T_i} F_i(y_{i,t-1})^2 &\to_d \quad L_i(1,0)\int_{-\infty}^{\infty} F_i(s)^2 ds \end{aligned}$$

Since $\triangle y_{i,t-k},\ k = 1,\ldots,P_i$ and $w_{i,t-j}\ j = 1,\ldots,Q_i$ are stationary, we also have

$$\begin{aligned} T_i^{-3/4}\sum_{t=1}^{T_i} F_i(y_{i,t-1})\triangle y_{i,t-k} &\to_p \quad 0, \quad \text{for all } k = 1,\ldots,P_i \\ T_i^{-3/4}\sum_{t=1}^{T_i} F_i(y_{i,t-1}) w_{i,t-j} &\to_p \quad 0, \quad \text{for all } j = 1,\ldots,Q_i \end{aligned}$$

due to the asymptotic orthogonality between stationary variables and integrable transformations of integrated processes established in Lemma 5 (e) of Chang, Park and Phillips (2001).

Using (11) and (13), we may write τ_i defined in (12) as

$$\tau_i = \frac{B_{T_i}^{-1}A_{T_i}}{(\hat\sigma_i^2 B_{T_i}^{-2}C_{T_i})^{1/2}} = \frac{A_{T_i}}{\hat\sigma_i C_{T_i}^{1/2}} = \frac{T_i^{-1/4}\displaystyle\sum_{t=1}^{T_i} F_i(y_{i,t-1})\varepsilon_{it}}{\hat\sigma_i\left(T_i^{-1/2}\displaystyle\sum_{t=1}^{T_i} F_i(y_{i,t-1})^2\right)^{1/2}} \ + \ o_p(1)$$

Now the state result follows immediately. ∎

Proof of Lemma 3.2 The asymptotic independence of $\tau_1,\ldots,\tau_M$ follows if we show that τ_i and τ_j are asymptotically orthogonal for all $i,j = 1,\ldots,M$. The proof goes exactly the same as that in Chang (2000), except for the pairs τ_i and τ_j for which the corresponding cross-sectional units (y_{it}) and (y_{jt}) are cointegrated. Note that Assumption 4.1 in Chang (2000) holds under our Assumption 3.2. Therefore, we set i and j, and assume (y_{it}) and (y_{jt}) are cointegrated

To establish the asymptotic orthogonality of τ_i and τ_j, it suffices to show that

$$\sqrt[4]{T_i T_j} \int_0^1 F_i(\sqrt{T_i} U_{iT_i}(r)) F_j(\sqrt{T_j} U_{jT_j}(r)) dr \to_p 0 \tag{22}$$

See Chang (2000) for details. As shown in Chang, Park and Phillips (2001), we have

$$\begin{aligned} & \frac{T_1}{\log T_1} \int_0^1 F_i(\sqrt{T_1} U_{iT_i}(r)) F_j(\sqrt{T_1} U_{jT_j}(r)) dr \\ = \; & \frac{T_1}{\log T_1} \int_0^1 F_i(\sqrt{T_1} U_i(r)) F_j(\sqrt{T_1} U_j(r)) dr + o_p(1) \end{aligned}$$

However, due to our convention in (8) made on scale adjustment, we may assume that

$$\begin{aligned} & \int_0^1 F_i(\sqrt{T_i} U_{iT_i}(r)) F_j(\sqrt{T_j} U_{jT_j}(r)) dr \\ = \; & \int_0^1 F_i(\sqrt{T_1} U_{iT_i}(r)) F_j(\sqrt{T_1} U_{jT_j}(r)) dr \end{aligned}$$

and that

$$U_i = U_j$$

Moreover, as shown in, e.g., Revuz and Yor (1994), we have

$$\int_0^1 (F_i F_j)(\sqrt{T_1} U_i(r)) = O_p(T_1^{-3/4})$$

since F_i and F_j are assumed to be orthogonal.

We may assume without loss of generality that the numeraire unit has the largest number of observations, i.e., $T_1 \geq T_i$ for all i. Then we have

$$\begin{aligned} & \sqrt[4]{T_i T_j} \int_0^1 F_i(\sqrt{T_i} U_{iT_i}(r)) F_j(\sqrt{T_j} U_{jT_j}(r)) dr \\ = \; & \sqrt[4]{T_i T_j} \int_0^1 F_i(\sqrt{T_1} U_{iT_i}(r)) F_j(\sqrt{T_1} U_{jT_j}(r)) dr \\ = \; & \sqrt[4]{T_i T_j} \int_0^1 (F_i F_j)(\sqrt{T_1} U_i(r)) dr + o_p\left(\frac{T_i^{1/4} T_j^{1/4} \log T_1}{T_1} \right) \\ = \; & O_p\left(T_j^{1/4} / T_i^{1/2} \right) \\ = \; & O_p\left(T_{\max}^{1/4} / T_{\min}^{1/2} \right) \end{aligned}$$

since T_1 is set at $T_{\max}$. Now (22) follows immediately given Assumption 3.2. ∎

Proof of Theorem 3.3 We first consider the case $M = N$. The statistic S has standard normal limiting distribution, and therefore, the stated result follows immediately. For the statistic $S_{\min}$, we note that

$$\begin{aligned} \lim_{T \to \infty} \mathbf{P} \{ S_{\min} \leq x \} &= \lim_{T \to \infty} \mathbf{P} \left\{ \min_{1 \leq i \leq N} \tau_i \leq x \right\} \\ &= 1 - \prod_{i=1}^{N} \lim_{T_i \to \infty} \mathbf{P} \{ \tau_i > x \} \\ &= 1 - (1 - \Phi(x))^N \end{aligned}$$

since τ_i's are asymptotically independent normals.

For the case $1 \leq M \leq N$, we have

$$\begin{aligned} \lim_{T\to\infty} \mathbf{P}\left\{S_{\max} \leq x\right\} &= \lim_{T\to\infty} \mathbf{P}\left\{\max_{1\leq i\leq N} \tau_i \leq x\right\} \\ &= \prod_{i=1}^{N} \lim_{T_i\to\infty} \mathbf{P}\{\tau_i \leq x\} \\ &= \Phi(x)^M \end{aligned}$$

Note that for $i = 1, \ldots, M$

$$\lim_{T_i\to\infty} \mathbf{P}\{\tau_i \leq x\} = \Phi(x)$$

and for $i = M+1, \ldots, N$

$$\lim_{T_i\to\infty} \mathbf{P}\{\tau_i \leq x\} = 1$$

since $\tau_i \to_p -\infty$ as $T_i \to \infty$ in this case. Therefore,

$$\lim_{T\to\infty} \mathbf{P}\left\{S_{\max} \leq x_M(\lambda)\right\} = \Phi(x_M(\lambda))^M = \lambda$$

and

$$\lim_{T\to\infty} \mathbf{P}\left\{S_{\max} \leq x(\lambda)\right\} = \Phi(x(\lambda))^M = \lambda\Phi(x(\lambda))^{M-1} \leq \lambda$$

as was to be shown. The consistency of the tests then follows immediately from the result in Lemma 3.1. ∎

References

Baltagi, B.H. and C. Kao (2000). "Nonstationary panels, cointegration in panels and dynamic panels: A survey," mimeographed, Department of Economics, Texas A& M University.

Banerjee, A. (1999). "Panel data unit roots and cointegration: An Overview," *Oxford Bulletin of Economics & Statistics* **61**: 607-629.

Chang, Y. (1999). "Bootstrap unit root tests in panels with cross-sectional dependency," Rice University, mimeographed.

Chang, Y. (2000). " Nonlinear IV unit root tests in panels with cross-sectional dependency," forthcoming in *Journal of Econometrics.*

Chang, Y. and J.Y. Park (1999). "Nonstationary index models," mimeographed, Department of Economics, Rice University.

Chang, Y., J.Y. Park and P.C.B. Phillips (2001). "Nonlinear econometric models with cointegrated and deterministically trending regressors," *Econometrics Journal*, 4, 1-36.

Choi, I. (2001a). "Unit root tests for panel data," *Journal of International Money and Finace*, 20, 219-247.

Choi, I. (2001b). "Unit root tests for cross-sectionally correlated panels," mimeographed, Kukmin University.

Frankel, J. and A. Rose (1996). “A panel project on purchasing power parity: Mean reversion within and between countries,” *Journal of International Economics*, 40, 209-224.

Hansen, B.E. (1995). “Rethinking the univariate approach to unit root testing: Using covariates to increase power,” *Econometric Theory*, 11, 1148-1171.

Im, K.S., M.H. Pesaran and Y. Shin (1997). “Testing for unit roots in heterogeneous panels,” mimeographed.

Levin, A., C.F. Lin and C.S. Chu (1997). “Unit root tests in panel data: Asymptotic and finite sample properties,” University of California, San Diego, mimeographed.

Kasahara, Y. and S. Kotani (1979). “On limit processes for a class of additive functionals of recurrent diffusion processes,” *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **49**: 133-153.

MacDonald, R. (1996). “Panel unit root tests and real exchange rates,” *Economics Letters*, 50, 7-11.

Maddala, G.S. and S. Wu (1999). “A comparative study of unit root tests with panel data and a new simple test: Evidence from simulations and bootstrap,” *Oxford Bulletin of Economics and Statistics*, 61, 631-.

O’Connell, P.G.J. (1998). “The overevaluation of purchasing power parity,” *Journal of International Economics*, 44, 1-20.

Oh, K.Y. (1996). “Purchasing power parity and unit root tests using panel data,” *Journal of International Money and Finance*, 15, 405-418.

Papell, D.H. (1997). “Searching for stationarity: Purchasing power parity under the current float,” *Journal of International Economics*, 43, 313-323.

Park, J.Y. and P.C.B. Phillips (1999). “Asymptotics for nonlinear transformations of integrated time series,” *Econometric Theory*, **15**, 269-298.

Park, J.Y. and P.C.B. Phillips (2001). “Nonlinear regressions with integrated time series,” *Econometrica*, 69, 117-162.

Phillips, P.C.B. and V. Solo (1992). “Asymptotics for linear processes,” *Annals of Statistics*, 20, 971-1001.

Phillips, P.C.B. and D. Sul (2001). “Dynamic panel estimation and homogeneity testing under cross section dependence,” mimeographed, University of Auckland.

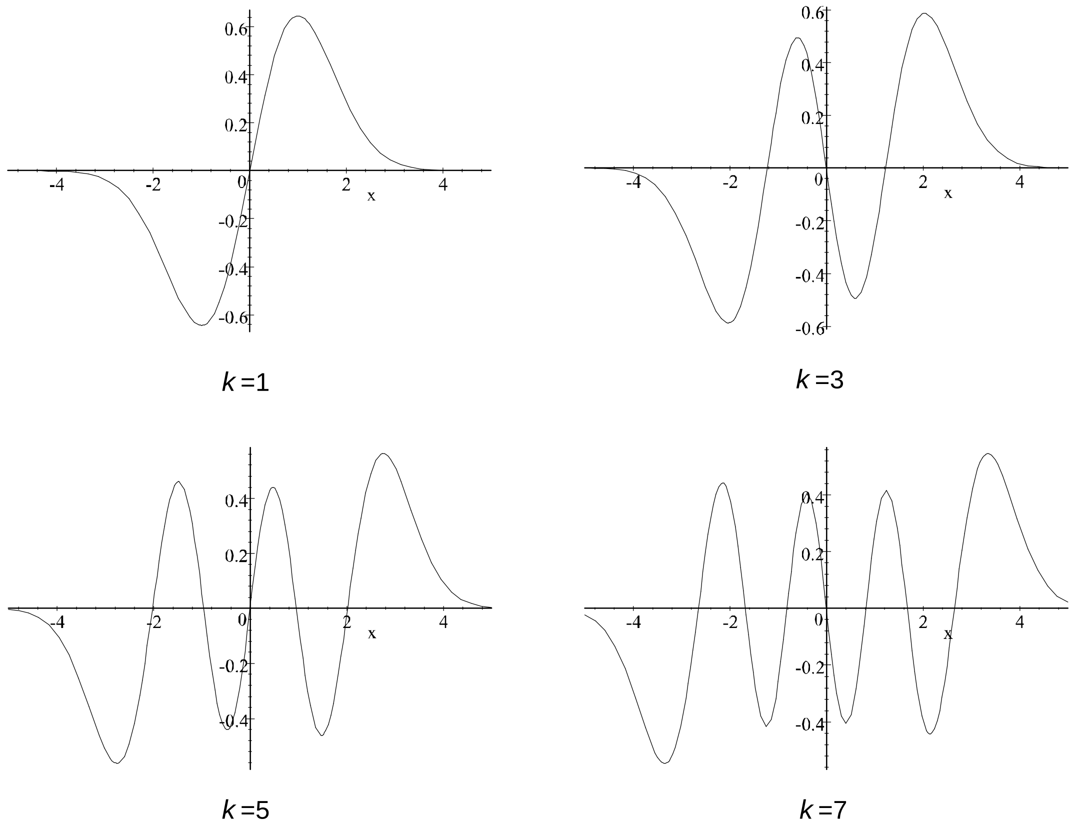

Figure 1: The Hermite Functions: $\psi_k(x) = (2^k k! \sqrt{\pi})^{-1/2} H_k(x) e^{-x^2/2}$

Table A: Critical Values $x_M(\lambda)$ for Maximum Statistic $S_{\max}$

M	1%	5%	10%	M	1%	5%	10%	M	1%	5%	10%
2	-1.282	-0.760	-0.478	42	1.260	1.484	1.613	82	1.602	1.801	1.916
3	-0.788	-0.336	-0.090	43	1.273	1.496	1.624	83	1.607	1.806	1.921
4	-0.478	-0.068	0.157	44	1.285	1.508	1.635	84	1.613	1.811	1.926
5	-0.258	0.124	0.334	45	1.297	1.519	1.646	85	1.619	1.817	1.931
6	-0.090	0.271	0.471	46	1.309	1.530	1.656	86	1.624	1.822	1.936
7	0.045	0.390	0.582	47	1.320	1.540	1.666	87	1.630	1.827	1.941
8	0.157	0.489	0.674	48	1.332	1.551	1.676	88	1.635	1.832	1.946
9	0.252	0.574	0.753	49	1.343	1.561	1.686	89	1.641	1.837	1.951
10	0.334	0.647	0.822	50	1.353	1.570	1.695	90	1.646	1.842	1.956
11	0.407	0.711	0.882	51	1.364	1.580	1.704	91	1.651	1.847	1.960
12	0.471	0.769	0.936	52	1.374	1.589	1.713	92	1.656	1.852	1.965
13	0.529	0.821	0.985	53	1.384	1.599	1.722	93	1.661	1.856	1.969
14	0.582	0.868	1.029	54	1.393	1.608	1.731	94	1.666	1.861	1.974
15	0.630	0.911	1.070	55	1.403	1.616	1.739	95	1.671	1.866	1.978
16	0.674	0.951	1.108	56	1.412	1.625	1.747	96	1.676	1.870	1.983
17	0.715	0.988	1.142	57	1.421	1.633	1.755	97	1.681	1.875	1.987
18	0.753	1.022	1.175	58	1.430	1.642	1.763	98	1.686	1.879	1.991
19	0.788	1.054	1.205	59	1.439	1.650	1.771	99	1.691	1.884	1.996
20	0.822	1.084	1.233	60	1.447	1.658	1.779	100	1.695	1.888	2.000
21	0.853	1.113	1.260	61	1.456	1.665	1.786	101	1.700	1.892	2.004
22	0.882	1.139	1.285	62	1.464	1.673	1.793	102	1.704	1.897	2.008
23	0.910	1.164	1.309	63	1.472	1.680	1.801	103	1.709	1.901	2.012
24	0.936	1.188	1.332	64	1.480	1.688	1.808	104	1.713	1.905	2.016
25	0.961	1.211	1.353	65	1.488	1.695	1.814	105	1.718	1.909	2.020
26	0.985	1.233	1.374	66	1.495	1.702	1.821	106	1.722	1.913	2.024
27	1.008	1.253	1.393	67	1.503	1.709	1.828	107	1.727	1.917	2.028
28	1.029	1.273	1.412	68	1.510	1.716	1.834	108	1.731	1.921	2.032
29	1.050	1.292	1.430	69	1.518	1.723	1.841	109	1.735	1.925	2.035
30	1.070	1.310	1.447	70	1.525	1.729	1.847	110	1.739	1.929	2.039
31	1.089	1.328	1.464	71	1.532	1.736	1.853	111	1.743	1.933	2.043
32	1.108	1.345	1.480	72	1.539	1.742	1.860	112	1.747	1.937	2.047
33	1.125	1.361	1.495	73	1.545	1.748	1.866	113	1.751	1.940	2.050
34	1.142	1.376	1.510	74	1.552	1.754	1.872	114	1.755	1.944	2.054
35	1.159	1.392	1.525	75	1.559	1.761	1.877	115	1.759	1.948	2.057
36	1.175	1.406	1.539	76	1.565	1.767	1.883	116	1.763	1.952	2.061
37	1.190	1.420	1.552	77	1.571	1.772	1.889	117	1.767	1.955	2.064
38	1.205	1.434	1.565	78	1.578	1.778	1.894	118	1.771	1.959	2.068
39	1.219	1.447	1.578	79	1.584	1.784	1.900	119	1.775	1.962	2.071
40	1.233	1.460	1.590	80	1.590	1.790	1.905	120	1.779	1.966	2.075
41	1.247	1.472	1.602	81	1.596	1.795	1.911	121	1.782	1.969	2.078

Table B: Critical Values $y_N(\lambda)$ for Minimum Statistic $S_{\min}$

N	1%	5%	10%	N	1%	5%	10%	N	1%	5%	10%
2	-2.575	-1.955	-1.632	42	-3.492	-3.031	-2.806	82	-3.667	-3.227	-3.015
3	-2.712	-2.121	-1.818	43	-3.499	-3.038	-2.814	83	-3.670	-3.231	-3.019
4	-2.806	-2.234	-1.943	44	-3.505	-3.045	-2.821	84	-3.673	-3.234	-3.022
5	-2.877	-2.319	-2.036	45	-3.511	-3.051	-2.828	85	-3.676	-3.237	-3.026
6	-2.934	-2.386	-2.111	46	-3.517	-3.058	-2.835	86	-3.679	-3.241	-3.030
7	-2.981	-2.442	-2.172	47	-3.522	-3.064	-2.842	87	-3.682	-3.244	-3.033
8	-3.022	-2.490	-2.224	48	-3.528	-3.071	-2.849	88	-3.685	-3.247	-3.037
9	-3.057	-2.531	-2.269	49	-3.533	-3.077	-2.856	89	-3.688	-3.250	-3.040
10	-3.089	-2.568	-2.309	50	-3.539	-3.083	-2.862	90	-3.691	-3.254	-3.043
11	-3.117	-2.601	-2.344	51	-3.544	-3.089	-2.868	91	-3.694	-3.257	-3.047
12	-3.143	-2.630	-2.376	52	-3.549	-3.094	-2.874	92	-3.697	-3.260	-3.050
13	-3.166	-2.657	-2.406	53	-3.554	-3.100	-2.880	93	-3.699	-3.263	-3.053
14	-3.187	-2.682	-2.432	54	-3.559	-3.106	-2.886	94	-3.702	-3.266	-3.056
15	-3.207	-2.705	-2.457	55	-3.564	-3.111	-2.892	95	-3.705	-3.269	-3.060
16	-3.226	-2.726	-2.480	56	-3.569	-3.116	-2.898	96	-3.707	-3.272	-3.063
17	-3.243	-2.746	-2.502	57	-3.573	-3.122	-2.903	97	-3.710	-3.275	-3.066
18	-3.259	-2.765	-2.522	58	-3.578	-3.127	-2.909	98	-3.713	-3.278	-3.069
19	-3.275	-2.783	-2.541	59	-3.582	-3.132	-2.914	99	-3.715	-3.281	-3.072
20	-3.289	-2.799	-2.559	60	-3.587	-3.137	-2.919	100	-3.718	-3.283	-3.075
21	-3.303	-2.815	-2.576	61	-3.591	-3.141	-2.924	101	-3.720	-3.286	-3.078
22	-3.316	-2.830	-2.592	62	-3.595	-3.146	-2.929	102	-3.723	-3.289	-3.081
23	-3.328	-2.844	-2.607	63	-3.599	-3.151	-2.934	103	-3.725	-3.292	-3.084
24	-3.340	-2.858	-2.621	64	-3.603	-3.155	-2.939	104	-3.728	-3.294	-3.087
25	-3.351	-2.870	-2.635	65	-3.607	-3.160	-2.944	105	-3.730	-3.297	-3.089
26	-3.362	-2.883	-2.648	66	-3.611	-3.164	-2.949	106	-3.732	-3.300	-3.092
27	-3.373	-2.895	-2.661	67	-3.615	-3.169	-2.953	107	-3.735	-3.302	-3.095
28	-3.383	-2.906	-2.673	68	-3.619	-3.173	-2.958	108	-3.737	-3.305	-3.098
29	-3.392	-2.917	-2.685	69	-3.623	-3.177	-2.962	109	-3.739	-3.308	-3.100
30	-3.402	-2.928	-2.696	70	-3.627	-3.182	-2.967	110	-3.742	-3.310	-3.103
31	-3.411	-2.938	-2.707	71	-3.630	-3.186	-2.971	111	-3.744	-3.313	-3.106
32	-3.419	-2.948	-2.718	72	-3.634	-3.190	-2.976	112	-3.746	-3.315	-3.108
33	-3.428	-2.957	-2.728	73	-3.637	-3.194	-2.980	113	-3.749	-3.318	-3.111
34	-3.436	-2.966	-2.738	74	-3.641	-3.198	-2.984	114	-3.751	-3.320	-3.114
35	-3.444	-2.975	-2.747	75	-3.644	-3.201	-2.988	115	-3.753	-3.323	-3.116
36	-3.451	-2.984	-2.756	76	-3.648	-3.205	-2.992	116	-3.755	-3.325	-3.119
37	-3.459	-2.992	-2.765	77	-3.651	-3.209	-2.996	117	-3.757	-3.327	-3.121
38	-3.466	-3.000	-2.774	78	-3.654	-3.213	-3.000	118	-3.759	-3.330	-3.124
39	-3.473	-3.008	-2.782	79	-3.658	-3.216	-3.004	119	-3.761	-3.332	-3.126
40	-3.479	-3.016	-2.791	80	-3.661	-3.220	-3.008	120	-3.764	-3.334	-3.129
41	-3.486	-3.023	-2.799	81	-3.664	-3.224	-3.011	121	-3.766	-3.337	-3.131